Home Advantage in Football; does distance play a role?

W.A.L. van Hattum

Abstract:
The home team advantage is one of the most interesting phenomena in sports. This research has looked at 26,468 games played between 2002 and 2014 at the four highest levels of professional English football. Individual match analysis using an ordered probit model shows that travel distance has a positive and significant effect on the home team advantage. An increase of 100 kilometers traveled distance, by the away team, increased the home team points by 0.02 points, c.p. Actual effects are six times larger than the incorporated expected effect by bookmakers, concerning traveled distance. This opens up the possibility for profitable betting strategies and contradicts that the betting market functions under the efficient market hypothesis.

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Master Thesis Urban, Port & Transport Economics

W.A.L. van Hattum
331510wh

Supervisor: G. Mingardo
First Reader: E.Braun

Date final version: August 24, 2017

Contents

1. Introduction

Besides the social importance of sports, the main reason that sports economics is a popular research subject is that it functions as a social laboratory in which *standard* economic theory could be easily tested. One of these theories is the physical location dilemma, where to practice business, to live or to invest heavily depends on the geographical location and regulation and is at the center of economic theory. To test this theory in sports we look at the effect location and travel distance has on match outcomes.

Home advantage is one of the most fascinating and widely discussed phenomena in sports. While the existence of the home advantage is well-known and not in dispute, the magnitude of the effect and the sources of home advantage are the center of discussion. One of the potential sources of home advantage is the distance traveled by away teams resulting is worn out players and a lower level of freshness (Schwartz & Barsky, 1977). As Steve Sands points out on the popular football website 11v11.com, which is home of The Association of Football Statisticians; “I still can’t help having a gut feeling that travel distance must be a factor when Carlisle United had to trek all the way down to Plymouth on a Tuesday evening!”.

Therefore the main question of the research is: “Does travel distance of away team’s affect the results in a football match?” Although, as will be showed in the theoretic framework, some researchers tried to answer this question already this research will use an extensive dataset and make use of modern up-to-date methods to finally put the discussion to rest.

In this research result of 26,468 matches, played between 2002 and 2014, at the four highest levels of professional English football, are examined. We found a home team advantage of 58.6% which is in line with previous findings. Using an ordered probit model we further show that travel distance has a positive and significant effect at the 1% level. An increase of 100 kilometers traveled distance by the away team leads to an increase of the home team points by 0.02 points. To the best of our knowledge, this is the first attempt to calculate the direct effect traveled distance has on the expected number of points. As in standard economics, the physical location does affect the outcome.

A novelty of this research is that the possible influence of travel fatigue affecting home advantage is combined with bookmaker expectations. We find that arbitrage is possible due to an incorrect incorporation of the effect traveled distance has on the actual result. Actual effects are six times larger than the incorporated expected effect by bookmakers, concerning traveled distance.

The following section of this research will contain a theoretical framework on which this research is built. Section three will describe the data used in this research while section four handles the methodology of this research. Section five and six will show the results of this research, conclude and discusses the implications of these results. This research will be finalized with some final remarks and points for further research.

2. Theoretical framework

This section of this research will contain the theoretical framework upon which this research is built. This section will start with a brief discussion about the importance of academic literature concerning sports economics, followed by the introduction of the home advantage concept. An overview of the available literature concerning the home advantage in sports in general and football in specific will be next. After focusing on the impact of location and geographical distance in light of the home advantage, the first hypothesis is formed. To conclude this section the concept of home team advantage is linked to bookmaker's expectations introducing the second hypothesis.

2.1 The economics of sport

In 1998 Chris Gatton wrote in his article "The economic importance of modern sport" that sport was one of the fastest growing sectors in the leisure business (Gratton, 1998). But estimations of revenue streams of professional football clubs show that the sports industry is still fairly small compared to other businesses. For example, all revenue streams of professional football clubs combined account for only half of the revenue stream of a multinational company like Heineken (Deloitte, 2017). Still an extensive stream of academic literature concerning sport is published by major journals. Why? Besides the social importance of sports the main reason for this is that sport economics function as a social laboratory in which *standard* economics theory could be easily tested. Thaler and Ziemba even argued in their paper that the sports betting industry is the better suited to check market efficiencies and rationality than the stock market (Thaler & Ziemba, 1988). Due to clear rules of the game, high incentives and a lot of media attention, the industry functions as a perfect place for research in different aspects of modern economy (e.g. labor economics, behavioral economics, and industrial economics).

One of the economic theories that could be tested in a sport environment is the location choice decision. The physical location dilemma, where to practice business, to live or to invest heavily depends on the geographical location and regulation and is at the center of economic theory. For instance, transportation costs and trade volume depend on infrastructure and

geographical location and some locations are worse than others resulting in a geographical disadvantage (Limao & Venables, 2001). But is this also applicable to the sports industry?

2.2 Home advantage in sports

Home advantage is one of the most fascinating and widely discussed phenomena in sports. It describes the advantage a home team has over his opponent. This advantage can be attributed to beneficial physiological factors for the home team or detrimental factors that influence the away team. In all professional team sports, the team playing at home derives an advantage (Pollard R. , 1986).

Reasons for this phenomena were already subject to research in the late seventies by Schwartz & Barsky. The authors suggest three possible sources for the home advantage. First home team players have a greater intimacy and knowledge of the home pitch and ballparks [baseball]. This familiarity with the conditions would lead to an edge over their opponents. The second source of the home advantage could be that the home team is less fatigue and could make use of better facilities than away teams. Travel is tiring, visitor teams maybe were worn out from traveling which would influence their performance negatively. The third factor that might influence the home advantage is the local support by fans. The moral support would increase performance of home team players (Schwartz & Barsky, 1977). The authors were also among the first researchers to quantify the home advantage. Their findings showed that the home advantage varied from sport to sport and indoor sports were influenced more than outdoor sports. Team strength also played a role in the magnitude of the home advantage. The home advantage was biggest when the home team was strong and the away team weak (Schwartz & Barsky, 1977).

Around the same period John Edwards in his contribution to the book *Sport, Games and Play* showed that over a period of three years American football teams won 54.4% of all their home games (Edwards, 1979). Edwards conclude that his findings were in line with the research of Schwartz and Barsky. Edwards found similar results for other sport such as baseball and basketball. Despite the home advantage occurred in each sport is varied in percentage and no pattern could be seen. For instance, the home advantage in the second part of the season was

bigger than the first part for professional Football teams while this trend reversed for college teams (Edwards, 1979).

Research published in the Journal of Applied Statistics done by Stefani and Clarke in 1992 showed that the home advantage was beside (college) football also present in hockey, soccer and Australian Rules football. The authors concluded that the home advantage varied in magnitude from sport to sport and could be due to different reasons (Stefani & Clarke, 1992). More recent research supports earlier findings. Swartz and Arce find a home advantage for teams in the National Hockey League [NHL] of 54.5% while NBA teams have a 60.5% home advantage. The author's further state that the magnitude of home advantage has declined over the years (Swartz & Arce, 2014). When looking at individual team results the home game advantage of NBA teams the Denver Nuggets and the Utah Jazz are the largest in their leagues. Both teams play at high altitude which suggests that playing at high altitude provides an edge in the home team advantage (Swartz & Arce, 2014).

2.3 Home advantage in football

As already stated in the previous part, Richard Pollard wrote in 1986 that in all professional team sports an advantage is derived by the team playing at home. Football is not an exception to this. On the contrary, in his research *Home advantage in soccer: a retrospective analysis* he argues that the home advantage in football is the biggest of all sports (Pollard R. , 1986). He quantifies the home advantage as the number of games won at home expressed as a percentage of all games played that season. If half of all wins of a team were at home, a fifty percent score, the home advantage would be none. This method is frequently used in American sports due to the absence of draws. In European football, a draw occurs almost a quarter of all times so Pollard adjusted his method and defined the home advantage as the number of points won at home with respect to the total number of point awarded. In his research on the First Division of the Football League in England from 1888 to 1984 he found a home advantage of 63.9% or higher for all different time periods (Pollard R. , 1986). He estimated that this advantage could be translated to an extra 0.6 goal for the home playing team every game. In addition to his research in 1986, Richard Pollard published an extensive research twenty years later in which he explored the existence of the home advantage in football around the word. Results show that the home advantage in Europe (1998 – 2004)

varied from 78.95% [Bosnia] to 48.87% [Andorra]. The Balkan countries all score high while countries in northern Europe have a smaller home advantage. In South America the Andean nations scored above average while the southerly countries scored lower. Countries in the rest of the world showed comparable figures to the European mean of home advantage (Pollard R. , 2006).

In his research Pollard used different explanations for the home advantage in football. Using the three sources of home advantage mentioned earlier by Schwartz and Barksy, he first hypothesized six possible explanations for the home advantage in football (Pollard R. , 1986). Adding a referee bias, the use of special tactics and physiological factors to the list of; familiarity, Travel fatigue and local crowd support (Schwartz & Barsky, 1977). Later in his literature review: *Home advantage in football: A current review of an unsolved Puzzle,* Pollard increases the list of potential sources of home advantage to nine categories including territoriality, rule factors and the interaction of causes (Pollard R. , 2008). He argues that from none of these categories a clear level of influence is set. Although the literature, since his first article in 1986, rapidly expanded "the puzzle is still unsolved".

Also other researchers found similar results and found evidence for the home advantage in Football. Stefani and Clarke investigated six nations (England, Germany, Italy, Norway, Spain and Switzerland during the seasons 1980-81 and 1981-82, and found a home advantage of 63.8% [1] (Stefani & Clarke, 1992). These results are in line with the results Pollard [63.9%] fond in 1986.

Similar numbers are found in research done by Clarke and Norman. With home teams in the English Football Leagues winning 48.7% of their games, while away teams only winning 24.6% of the games. In a window of ten years [1981 – 1991] this led up to a home advantage of 63.2% [2] (Clarke & Norman, 1995). In their research the authors also look at individual team efforts. They find no significant division effect but state that home advantage is different for different years. The most important discovery the authors make is that they find evidence for the fact

[1] Calculated based on method used in (Pollard R. , 1986) with data from (Stefani & Clarke, 1992) 48.5% home wins and 28.1% draws.
[2] Calculated based on method used in (Pollard R. , 1986) with data from (Clarke & Norman, 1995) 48.7% home wins and 24.6% away wins.

that home advantage plays a bigger role in match result in respect to goal difference (Clarke & Norman, 1995).

Contradictory to the findings of Clarke and Norman new research using end of season league tables done by Nevill et al. show different magnitudes of home advantage between different divisions. The authors find that in the higher divisions the home advantage [61.5% [3] in English Premier League] is higher compared to the lower divisions [51.0% [4] in Scottish Second division] (Nevill, Newell, & Gale, 1996). They argue that one of the possible reasons for this might be the relationship to crowd size, which is one of the sources of home advantage mentioned earlier in the research of Schwartz and Barksy.

The research by Nevill et al. also introduced a new angle to the home advantage discussion. While Pollard already in 1986 argued that one of the sources of home advantage could be a referee bias, no additional research was done yet. Nevill et al. showed that 62% of the sending-offs were contributed to the away team. The authors conclude that the home team fans either provoke the visiting players into more reckless challenges or influence the referee into believing that the away player had committed more fouls (Nevill, Newell, & Gale, 1996).

Also the article by Carmichael and Thomas focusses on in-game effect of the home advantage. They find that away teams are punished significantly more with yellow and red card than home teams (Carmichael & Thomas, 2005). The authors argue that the reason for this observation, in line with the research of Nevill et al., could be caused by a referee bias or due to tactics which is mentioned by Pollard as one of the potential sources of home advantage (Pollard R. , 2008). Other research done by Boyko et al. supports these initial findings and show that crowd size (indirectly) affect the home advantage through the referee bias (Boyko, Boyko, & Boyko, 2007).

[3] Calculated based on method used in (Pollard R. , 1986) with data from (Nevill, Newell, & Gale, 1996) 214 home wins, 130 draws and 118 away wins.
[4] Calculated based on method used in (Pollard R. , 1986) with data from (Nevill, Newell, & Gale, 1996) 105 home wins, 68 draws and 100 away wins.

2.4 Traveled distance away team

One of the possible sources for home advantage is travel fatigue (Schwartz & Barsky, 1977; Pollard R. , 1986; Pollard R. , 2008). But while academics and researchers come to a unanimous conclusion regarding overall home advantage, they disagree about the influence and magnitude of travel fatigue.

In 1992 Courneya and Carron published a literature review: *The home advantage in sport competitions: A literature review* in which they stated that after extensive literature research they concluded that "the effects of travel on the home advantage were minimal" (Courneya & Carron, 1992). Also recent analysis of Brendan Kent, data analyst for the Portland Timbers a professional football team in the USA finds no proof of travel distance affecting home advantage (Kent, 2015)."When control for general home advantage and the relative strength of two teams, the physical distance between two teams had no effect on results from a statistical standpoint" concludes Kent.

Contrary to these conclusions are the (partial) findings of among others Pollard and Clarke and Norman. But the first researchers who tried to quantify the effect of travelled distance by the away team are Snyder and Purdy. In their research published in the Sociology of Sport Journal they showed that teams had a home winning percentage of 58.8 [84.6] when visiting team traveled less [more] than 200 miles (Snyder & Purdy, 1985).

Pollard later found conflicting evidence for the effect of travel distance. He showed that the home advantage in European Cup games was higher than in national football league matches which could be caused by the greater distance between the two clubs. And while the home advantage was only 56.1% in local derbies in respect to around 64% in all other matches (Pollard R. , 1986), his results couldn't prove that travel distance played a vital role in the home advantage. A few years later Stefani and Clarke supported the findings of Pollard. Clubs had 15 percentage point more chance of winning their home games in European cup matches in comparison to national league matches. The shorter distance travel in national league matches was one of the possible causes (Stefani & Clarke, 1992).

In 1995 Clarke and Norman were the first to find a highly statistical significant correlation between distance traveled by the away team and home advantage (Clarke & Norman, 1995). The authors argue that his effect is reflected by the high score of home advantage by Plymouth (ranked 1), Exeter and Carlisle which are all geographically isolated. These findings are in line with the works of John Goddard. In his article published in the International Journal of Forecasting he finds that geographical distance plays a significant influence on match outcomes (Goddard J. , 2005) and in his co-work with Asimakopoulos they find a positive and significant correlation between distance traveled and the home team advantage at the 1% level (Goddard & Asimakopoulos, 2004).

The most exploration to travel distance and the effect of the home advantage is done by Pollard, Silva and Medeiros. In their research they found that the effect of distance traveled by the away teams in Brazil, after controlling for ability, would increase by 0.115 goal every 1000km. In Brazil the largest distance between two teams is 4172km. This would lead to an extra 0.479 goal for the home team (Pollard, Silva, & Medeiros, 2008). Looking at individual team results they find that Paysandú who play at Belem, have the highest home advantage of the league. A possible reason for this is that this is the most remote location with Fortaleza EC [around 1000km] as nearest opponent. These findings are in line with the previous studies mentioned above and are in line with other research in remote city's which hosted football clubs had a high home team advantage (Pollard & Seckin, 2007). These include Van [Turkey], Xanthi [Greece] and Korçë [Albania] which all score high home advantage percentages.

Also in other sports travel distance affect home advantage. Research has shown that team in the NCAA [College Basketball] that travel more than 150 miles for a match reduces their chances of winning by 33.6% in respect to a team who played in their home region (Clay, Clay, & Bro, 2014). A possible reason for this could be a higher level of fatigue due to the travel between time zones. Schechtman-Rook investigated result in the NFL and concluded that when team had to move east one or more time zones the home advantage of the home team raised from 58% to 65% (Schechtman-Rook, 2013).

Looking at the available literature the following hypothesis could be formed and a rejection of this hypotheses is expected considering the literature.

➔ H1: the distance traveled by the away teams in professional football has no effect on the home advantage.

2.5 Bookmaker's expectations

Literature shows that researchers are still unsure about the magnitude of the effect travel fatigue has on the home team advantage and thus on match results. But how do bookmakers incorporate these uncertainties than?

Bookmakers have to set their odds prior to the event or match and these odds represent the expectations of the bookmakers. It's the relationship between these odds they set for betting and the true probabilities of the different outcomes that is important. A difference in odds and true probabilities could potentially lead to arbitrage possibilities, which is undesirable for the bookmakers. If for instance bookmakers expected team A to win 50% of the time while the actual chance was 40%, betting on team B would be beneficiary for the gambler. The difficulty here is calculating or knowing the true probability because events happen only once and only one outcome is observed. A traditional economic theory that is connected to this problem is the efficient market hypothesis [EMH] by Eugene Fama. The EMH is a theory in financial economics that states that asset prices fully reflect all available information (Fama, 1970). In that light, bookmakers always incorporate every information that is available and their odds are correct, with the current knowledge. This view is supported by Thaler and Ziemba, who argue that betting markets are a better place to test market efficiencies and rationality (Thaler & Ziemba, 1988).

So is arbitrage impossible in sports betting? Recent research looked at this problem and a few interesting articles were published. To start, arbitrage is possible in the betting market as proven by Cain et al. In their research about UK football betting they show that betting tends to have a favorite-longshot bias. Gamblers prefer high payouts with small probabilities over small prizes with a higher chance. This preference offsets the betting market and therefore odds deviate from the true probabilities (Cain, Law, & Peel, 2000).

Secondly, Dixon and Coles build a statistical model in which past results predicted future matches. The authors used a Poisson regression model for these predictions and the model showed positive returns when used as the basis of a betting strategy (Dixon & Coles, 1997). Such a model was also used by Goddard and Asimakopoulos. In their article published in the Journal of Forecasting they concluded that betting markets were not efficient. Their regression based test indicated that their forecasting model contained information about match results which was not incorporated into the bookmakers odd's (Goddard & Asimakopoulos, 2004). An optimal betting strategy based on their model (under certain restrictions) could yield an 8% margin.

In conclusion Goddard stated that although having an edge, due to variation, results could vary a lot and "betting informed by statistical analysis does not represent a quick or easy route toward sudden wealth and fortune" (Goddard J. , 2006). Therefore the second hypothesis of the research is:

- ➔ H2: There is no arbitrage possibility at bookmakers betting market regarding travel distance in football matches.

In the following sections these two hypotheses will be tested. The next section will describe the dataset used in this research and explains the variables used to examine this dataset.

3. Data

The dataset compiled for this research consists of thirteen years of professional English football. From season 2001-02 [referred to as season 2002] to season 2013-14 [season 2014] results of all matches played in the four highest divisions[5] of professional English football association are out together into one database. Therefore, the database consists of 26,468 [6] observations regarding English football. In those years, 109 professional football clubs [see Appendix A for a list of all teams] were active on those levels thus are included in this dataset. Results are checked for validity by using multiple sources [7] for match data. Individual match data includes; date, home team, away team, full-time result, half-time result, information about shots and cards for both teams and the appointed referee.

3.1 Dependent variables

Match results

As shown in Figure 1 the home team won 44% percent of all their games and lost 29% percentage. Translating this into the home advantage (Pollard R. , 1986) as referred to in the theoretical framework is for all divisions together 58.6% [63.9 (Pollard R. , 1986); 63.8 (Stefani & Clarke, 1992); 63.2 (Clarke & Norman, 1995)]. The home advantage is more dominant in the higher leagues but the variance between the levels is little, with the highest percentage being 60.3% [Premier League] and the lowest 57.3% [League Two]. Although the difference is very small this is in line with the findings of Nevill et al. which also showed that home advantage was more prominent in the higher levels of professional football (Nevill, Newell, & Gale, 1996).

	Win Percentage			Points		Goals		Yellow Cards		Red Cards	
Division	Home	Away	Draw	Home	Away	Home	Away	Home	Away	Home	Away
Premier League	0.466	0.277	0.256	1.655	1.088	1.527	1.125	1.332	1.761	0.066	0.099
Championship	0.441	0.283	0.276	1.600	1.124	1.464	1.125	1.284	1.675	0.069	0.108
League One	0.442	0.286	0.272	1.599	1.129	1.476	1.146	1.243	1.650	0.073	0.113
League Two	0.426	0.294	0.280	1.559	1.161	1.425	1.117	1.220	1.593	0.074	0.116
Total	0.442	0.286	0.272	1.599	1.129	1.469	1.129	1.264	1.662	0.071	0.110

Figure 1

[5] 2001 to 2003: Premier League, Division 1, Division 2 and Division 3
From 2003 onwards: Premier League, Championship, League One and League Two.
[6] (13 x 20 x 19) + (13 x 24 x 23 x 3) = 26,468
[7] www.*football-data.co.uk* – *www.soccerway.nl* – *www.soccerbase.com*

When looking at scored goals [Figure 1], the home team averaged 1.469 goals per match while away teams only managed to score 1.129 goals per match. Resulting in a home advantage of 0.340 [0.6 (Pollard R. , 1986)] goals per match. Comparing these results to earlier research, it is safe to say that the home advantage has decreased, in results as well as goal difference, over time. This is in line with the conclusion from Swartz and Arce who stated that home advantage declined over the year's (Swartz & Arce, 2014).

This dataset also supports the findings of earlier research that away teams are punished more often with yellow and red cards (Nevill, Newell, & Gale, 1996; Carmichael & Thomas, 2005). As shown in Figure 1 away teams receive almost 30% more yellow cards across all divisions. For red cards, this number is even higher, 55%. Interesting is the fact that more yellow cards are awarded in the higher divisions while red cards are awarded more often in the lower leagues.

If we look at the individual clubs, the home advantage varies from 50.7% to 68.1%. Figure 2 shows the top five and bottom five teams regarding the home advantage.

Ranking	Team	Home Points	Away Points	Home Advantage
1	Boston United FC	1.670	0.783	68.09%
2	Fulham FC	1.587	0.785	66.89%
3	Halifax Town FC	1.043	0.522	66.67%
4	Gillingham FC	1.706	0.940	64.48%
5	Rushden & Diamonds FC	1.661	0.922	64.31%

105	AFC Wimbledon	1.275	1.087	53.99%
106	Milton Keynes Dons FC	1.678	1.435	53.91%
107	Hereford United FC	1.290	1.123	53.45%
108	Chester FC	1.087	1.026	51.44%
109	Fleetwood Town FC	1.500	1.457	50.74%

Figure 2

Bookmaker's expectations

To calculate the bookmaker's expectations this research makes use of the odd's set by bookmakers prior to the match. Data is collected from the website football-data.co.uk and odds are taken every Friday [Tuesday] afternoon for weekend [midweek] games. Odds are set by the following bookmakers: Gamebookers, Interwetten, Ladbrokes, Sportingbet, William Hill, Bet365 and Bet&Win. If multiple bookmakers offered odds for the same match an average of all available odds were taken to ensure optimal results and correct for potential outliers.

To translate these odds into expectations the following calculations were made:

1. Convert odds into percentages.

$$Expected\ percentage\ of outcomes = \frac{1}{Odds\ Home\ Win} \quad , \quad \frac{1}{Odds\ Draw} \quad , \quad \frac{1}{Odds\ Away\ Win}$$

2. Margin of bookmaker is extracted.

$$Expected\ percentage\ of\ home\ win - margin = \frac{Home\ Win\ \%}{(Home\ Win\ \% + Draw\ \% + Away\ Win\ \%)}$$

3. Expected points are calculated for home and away teams.

$$Expected\ points\ for\ home\ team = (\ Home\ Win\ \% \times 3\ ) + (\ Draw\ \% \times 1\ )$$

Final bookmaker expectations for home and away teams are at first sight comparable to actual results, as can be seen in Figure 3. The expected home advantage 58.4% [8] is slightly lower than the actual home advantage 58.6% [9]. Important to notice is that the variation between the different divisions is fairly limited.

	Actual Points		Expected Points	
Division	Home	Away	Home	Away
Premier League	1.655	1.088	1.604	1.135
Championship	1.600	1.124	1.595	1.130
League One	1.599	1.129	1.590	1.137
League Two	1.559	1.161	1.582	1.143
Total	1.599	1.129	1.592	1.136

Figure 3

[8] 1.592 / (1.592 + 1.136) = 0.584
[9] 1.599 / (1.599 + 1.136) = 0.586

3.2 Independent variables

Distance

Our main independent variable of interest in this research is the actual distance between the home and away teams. Figure 4 shows all locations of the stadiums which are used in this database. For this research, the actual distance is set by the distance between both stadiums in miles. This number is later converted to the metric system. Using a VBA Macro and Microsoft's API for calculating distances all possible matches are calculated in a matrix. These distances are the fastest route between both stadia as of the moment the matrix was made [10]. An important note is that distance between Stadium x and y is not equal to the distance between Stadium y and x. The reason for this could be that some roads are only one-way street.

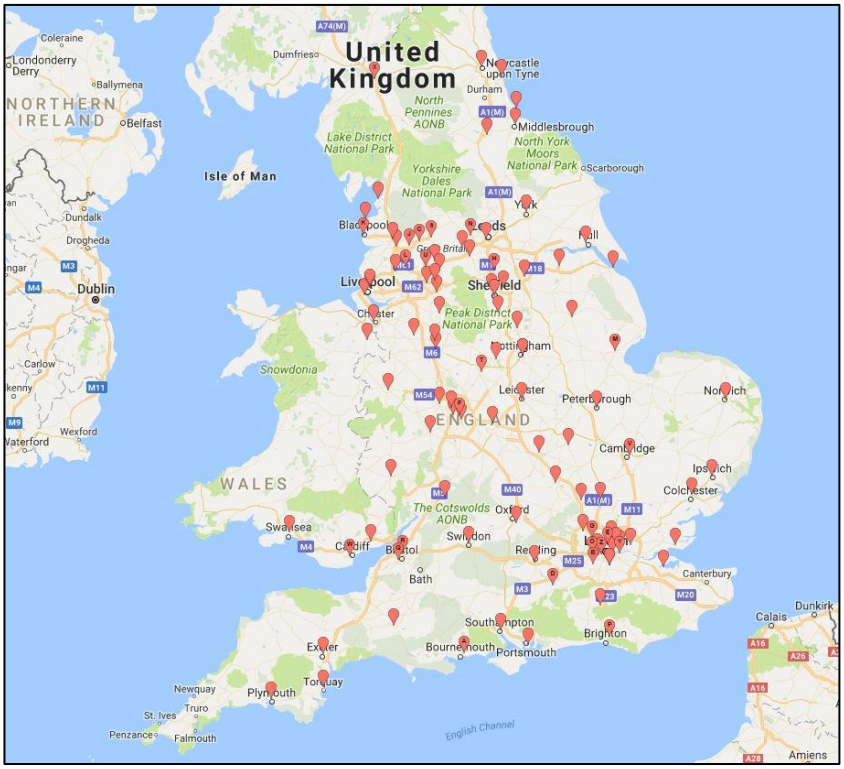

Figure 4

Overall the average distance between clubs in England is around 235 kilometer, with a standard variation of 125 kilometers. The actual distance between the clubs ranged from 0 kilometers (Crystal Palace vs Wimbledon, both teams played their home games at Selhurst Park Stadium) to 659 kilometers (Plymouth vs Newcastle).

[10] June 16, 2017.

Attendance

As suggested already by Barsky, Pollard and Carmichel one of the possible sources of the home advantage lies in the support of local fans (Schwartz & Barsky, 1977; Pollard R. , 2006; Carmichael & Thomas, 2005). Due to limited historical data of individual match attendances this research makes use of average home team attendance per year. Average attendance numbers are collected [11] for each team active in the first four divisions during the timespan of this research.

Average attendance was highest in Premier League [35,029] matches while League Two games had the lowest amount [4365] of spectators. At the individual club level, Figure 5, Manchester United had the highest attendance rates for all thirteen seasons in this database, averaging 72479 spectators each match. At the bottom end of this chart clubs like FC Dagenham & Redbridge, FC Halifax Town and FC Accrington Stanley only reaching an average match attendance each year of around 2,000 spectators per match. Overall the average attendance number grew from 13,707 in season 2002 to 14,554 in season 2014, whilst peaking in 2010 with an average number of spectators of 14,764.

Due to the high correlation with the division the game is played in, the attendance numbers in this database are relative to the average match attendance of the league the match is played in that year. This solves the potential problem of multicollinearity without leaving out one of the two variables. When interpreting the results this is important to keep in mind.

Ranking	Team	Average Attendance
1	Manchester United	72,479
2	Arsenal FC	51,579
3	Newcastle United FC	50,142
4	Liverpool FC	43,556
5	Manchester City FC	43,080

105	Barnet FC	2,272
106	Morcambe FC	2,254
107	Dagenham & Redbridge FC	2,161
108	Halifax Town FC	2,115
109	Accrington Stanley FC	1,778

Figure 5

[11] *http://www.european-football-statistics.co.uk/attn.htm*.

Other Control variables

One of the hardest things to control for is the ability of the football teams. In sports like baseball, the quality of the players is very well documented and is easily translated into statistical numbers (e.g. RBI, HR's). For football, this is much harder because more factors play a role and unobservable characteristics may dominate the outcome of the match. For this research, we, therefore, make use of rankings to determine the relative strength of each team. Two types of rankings are used to determine the relative strength of both teams in a fixture. The first one is the difference in standings of *last year rankings*. If a team was promoted or relegated from another division a mathematical correction was made to correct for the relative rankings. The second ranking is the difference in actual *current ranking* [12] on the moment the match is played. Using both indicators for relative strength this research controls for the difference in ability between the two teams. A third option, looking at the future ranking of that year would give a more accurate prediction of the strength. But due to a high correlation with the current ranking and the fact that at the moment the match is played the future ranking is impossible to know without foresight, we chose the first two rankings.

In addition to the two rankings, we the use difference in *market value* of the players between the teams to control for the same purpose, ability. This market value should depict the actual quality of the team as the market value of the players is based on their quality. Market value data is obtained from the website transfermarket.de and is incorporated in this database for matches from 2006 onwards.

Last but not least this research introduces a variable to control for the actual *form* of the team. This variable accounts for the difference in the number of points of both teams obtained in their last 3 [or 5] matches. For this variable, each team starts their *form* score at the beginning of the season at 0 and matches of the previous season are not taken in to account. In this research, the variable Form3 [Form5] is the difference between the form of both teams. The maximum [minimum] of this score is 9 [-9] for Form 3 and 15 [-15] for From5.

A correlation matrix between all variables is shown in Appendix B.

[12] Due the limited amount of data, the current rankings before a match are missing for the league one in 2002 and league two in 2002, 2003 and 2004.

4. Methodology

In this section, the methodology of this research is discussed. Using techniques from earlier research and new insights regarding this subject, the hypotheses posted in the theoretical framework of this research will be tested. Following Clarke and Norman, we use data of individual games (Clarke & Norman, 1995) and not league standings as Pollard often does to look at the determinants of the home team advantage. Looking at match level increases the accuracy of the results and will give more reliable results.

First, the method of testing the first hypothesis, effect of traveled distance on match outcome, will be explained. This is followed by the methodology of the second hypothesis which tests if the bookmaker's expectations are open to arbitrage possibilities.

4.1 Effect of travel distance

The outcome of a single football match is limited to three ordinal options, a win [3 points], a draw [1 point] and a draw [0 points]. This is different from the situation most American sports are in. These sports have to have a winner and play goes on till they have one, draws are therefore not a possible outcome of a match. In these sports standard probit or logit models are often used when looked at individual match data. But due to the fact that in football a third outcome is possible, standard probit or logit models are not applicable to this situation. Using an ordered probit model with three potential outcomes solves this problem. Therefore in this research, an ordered probit model is used to calculate the effects of the independent variables on the match outcome.

One of the most important assumptions of the ordered probit model is the parallel regressions assumption (Liao, 1994). The model assumes that the coefficients that describe the relationship between two ordinal levels are equal to those of the coefficient that describes the relationship between the other levels. An additional test in STATA is done to measure the approximate likelihood-ratio of equality of coefficients across response categories. A non-significant result suggests that the parallel regressions assumption is met. In our full model [prob. > Chi2 = 0.7160] this is the case and therefore this parallel regression assumption holds.

Another important assumption of the ordered probit model is that all errors have to be homoscedastic. For ordered probit models this assumption is far more important than for OLS regressions as heteroscedastic errors could bias the parameter estimates (Williams, 2008). However, there is no statistical test available at this moment to test homoscedasticity for ordered probit models. Therefore researchers often assume homoscedasticity in their models. For this reason, this research assumes homoscedastic errors in the full model.

The ordered probit model which is used in this research is set-up step-by-step to control for any signs of multicollinearity. Any sudden shifts of signs, big differences in the magnitude of coefficients and sudden loss/gain of significance could hint to multicollinearity. Because the high correlation [see correlation matrix Appendix B] between the variables Form3 and Form5 only one will be incorporated into the full model. Looking at the step-by-step model and correlation matrix, there are no concerns about multicollinearity in this model. In

Equation *1* the full model is mathematically expressed.

$$PR\ (Outcome\ match) = \beta_0 + \beta_1\, Distance + \beta_2\, Attendance + \beta_3\, Diff.last\ year\ rank + \beta_4\, Diff.current\ rank + \beta_5\, Diff.Form + \beta_6\, Diff.market\ value + \beta_7\, Division + \in$$

Equation 1

The results of this ordered probit model will show the chances of each outcome happening, summing up to 1. Changes in the independent variables would, therefore, change the chance of winning, drawing and losing and therefore the expected number of points of the home team.

To reject or accept the first hypothesis we look at β_1. If a positive or negative coefficient is found, which is significantly different from 0, we'll reject the hypothesis that there is no effect of traveled distance on the match outcome. If there is found to be a significant effect the partial effects on the three different outcomes have to be calculated in order to get the effect

on the total expected number of points. These results are further being dealt with in the next section.

4.2 Components of bookmaker's expectations

The second hypothesis in this research will be tested with a standard ordinary least squares regression. For this OLS regression, we use the same independent variables as used previously. The difference is that for this hypothesis we use the bookmaker's expected points as the dependent variable. Because this is a continuous variable we can use OLS regression.

As with any other OLS regression, certain assumptions have to hold in order to get unbiased and correct parameter estimations. The OLS regression in this research is linear in parameters, makes use of random sampling, there is no multicollinearity [see section 4.1] and the conditional mean is zero. A robust regression analysis is used to correct for potential outliers and heteroscedastic errors. Equation 2 shows the full robust regression analysis;

$$Expected\ Points\ Home\ Team = \beta_0 + \beta_1\, Distance + \beta_2\, Attendance + \beta_3\, Diff.last\ year\ rank + \beta_4\, Diff.current\ rank + \beta_5\, Diff.\ Form + \beta_6\, Diff.market\ value + \beta_7\, Division + \in$$

Equation 2

We test whether β_1 is positive/negative and significant. If a significant and positive or negative effect is found, we compare the effects with the results found regarding the first hypothesis. Overall, if the effect traveled distance has on the bookmaker's expectations is found to be larger or smaller than the actual effect found, arbitrage is possible.

5. Results

As discussed in the previous section of this research an ordered probit model is used to calculate the effects the independent variables have on match outcomes. These results are addressed in the first part of this section. The last part of this section will contain the results regarding the second hypothesis of this research. The implications of these results will be discussed in section 6 of this research.

Table 1 shows the results of the step-by-step ordered probit model. Models 1 to 5 adds a variable every step. Model 6 and 7 each introduce a form variable. Looking at the log likelihood and the Pseudo R-squared [$R^2 = 0.0281$ *vs* $R^2 = 0.0282$] the variable Form 5 is a better predictor of match outcome. Model 8 adds the dummy variables for the divisions using the Form 5 variable to reach the full model.

VARIABLES	(1) Actual Points	(2) Actual Points	(3) Actual Points	(4) Actual Points	(5) Actual Points	(6) Actual Points	(7) Actual Points	(8) Actual Points
Distance (100km)	0.0199***	0.0222***	0.0219***	0.0213***	0.0187***	0.0187***	0.0188***	0.0181***
	(0.00560)	(0.00562)	(0.00562)	(0.00590)	(0.00676)	(0.00676)	(0.00676)	(0.00678)
Attendance (1000x)		0.0238***	0.0195***	0.0144***	0.00901***	0.00898***	0.00891***	0.00916***
		(0.00106)	(0.00109)	(0.00112)	(0.00139)	(0.00139)	(0.00139)	(0.00140)
Last Year Rank			0.00976***	0.00734***	0.00490***	0.00488***	0.00486***	0.00483***
			(0.000625)	(0.000671)	(0.000786)	(0.000786)	(0.000786)	(0.000786)
Current Rank				0.0181***	0.0160***	0.0148***	0.0143***	0.0142***
				(0.000833)	(0.000958)	(0.00107)	(0.00114)	(0.00114)
Market Value					1.89e-09***	1.87e-09***	1.85e-09***	1.89e-09***
					(1.42e-10)	(1.43e-10)	(1.43e-10)	(1.43e-10)
Form 3 matches						0.00747**		
						(0.00304)		
Form 5 matches							0.00705***	0.00733***
							(0.00249)	(0.00249)
Championship								-0.0808***
								(0.0260)
League One								-0.0964***
								(0.0261)
League Two								-0.150***
								(0.0260)
Constant cut1	-0.520***	-0.523***	-0.527***	-0.535***	-0.534***	-0.535***	-0.535***	-0.627***
	(0.0154)	(0.0155)	(0.0155)	(0.0162)	(0.0187)	(0.0187)	(0.0187)	(0.0263)
Constant cut2	0.192***	0.198***	0.199***	0.198***	0.200***	0.199***	0.199***	0.108***
	(0.0152)	(0.0153)	(0.0153)	(0.0160)	(0.0185)	(0.0185)	(0.0185)	(0.0261)
Observations	26,468	26,468	26,468	24,260	18,324	18,324	18,324	18,324
Pseudo R2	0.000223	0.00943	0.0137	0.0234	0.0280	0.0281	0.0282	0.0290

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1

Overall we don't see any sudden shifts in signs, big differences in magnitude or significance. This, like already discussed in the methodology section, suggests that there is no problem concerning multicollinearity. With adding multiple variables, the number of observations drops due to the lack of availability of data. This is not a problem because more than 18,000 observations are still included in the full model.

The main independent variable of interest, the distance traveled by away teams, is found stable throughout the process of adding other variables. Distance traveled has a positive and significant effect, at the 1% level, on the match outcome. Due to the characteristics of the ordered probit model the magnitude of this effect is harder to determine and is not directly interpretable from Table 1. Using the marginal effects, these can be found in Table 2, this research concludes that every added 100 kilometers of travelling results in an increase of winning of 0.7%, decrease in drawing of 0.1% and a decrease in losing of 0.6%, c.p. Translating this into points, every 100 kilometer extra traveled by the away team will increase the number of points of the home team with 0.020, c.p. Given this result, we can, in line with the theoretical framework, reject the first hypothesis that traveled distance has no effect on match outcomes.

VARIABLES	Loss	Draw	Win	Points
Distance (100km)	-0.0061136	-0.0010125	0.0071261	0.0203658
Attendance (1000x)	-0.0030958	-0.0005127	0.0036085	0.0103128
Last Year Rank	-0.0016309	-0.0002701	0.0019011	0.0054332
Current Rank	-0.0047973	-0.0007945	0.0055918	0.0159809
Market Value	-6.38e-10	-1.06e-10	7.43e-10	2.13e-09
Form 5 matches	-0.0024777	-0.0004104	0.002888	0.0082536
Championship	0.027593	0.0041331	-0.0317261	-0.0910452
League One	0.0329739	0.0048387	-0.0378126	-0.1085991
League Two	0.0516256	0.0070365	-0.0586621	-0.1689498

Table 2

If we look at the other variables used in this research, results are often as expected. The variable attendance, as can be seen in Table 1, is found positive and significant at the 1% level. The magnitude of the effect an additional 1000 spectators (in relation to the mean of that year's average attendance on the level the match is played) to the football is an increase in 0.010 points to the home team, c.p. This is in line with the conclusions drawn by Schwartz and Barsky who argued that moral support was one of the drivers behind the home team advantage (Schwartz & Barsky, 1977).

As expected also ability plays a positive role in the match outcomes. Last year rank and current year rank both play a positive and significant role at the 1% level. An increase of the difference in current [last year] rankings in favor of the home team between both team will increase the home team points by 0.016 [0.005], c.p. An example of this would be that if the current leader of the league would play the number fifth of the current ladder they would be expected to get 0.048 [13] more points in a home game than if they played the second placed team of the current ladder, c.p. Also, the other variable that accounts for the difference in ability between both team, the difference in market value of both teams, is found to have a significant and positive effect at the 1% level. An increase in the difference of 1 million euro's in favor of the home team, would lead to an increase in 0.002 points to the home team, c.p.

When looking at the form variable results show that current form, based on the last five matches played, has a significant and positive effect on match outcome at the 1% level. An increase of one point in the difference of points accumulated in the last five matches would increase the expected home team points by 0.008, c.p. With a maximum of 15 points difference in form over the last 5 games, this could potentially increase the expected points of the home team by 0.124 [14] in the match.

A joint f-test shows that the dummy variable's used to correct for the level at which the game is played have found to have a significant effect on match outcome at the 1% level. A game played at the Championship level would decrease the expected points of the home team by

[13] Difference +3; 3 x 0.0159809 = 0.047943
[14] 15 x 0.008 = 0.123804

0.091 compared to if the matched would be played at Premier League level, c.p. For League One and League Two matches the decrease in expected home team points is even larger. If a match would be played in League One [League Two] the home team would be expected to get 0.109 [0.169] points less compared to if the match was played in the Premier League. These results are in line with the findings of Nevill et al., who concluded that the home team advantage decreased at lower levels of professional football (Nevill, Newell, & Gale, 1996). As discussed in section three of this research, the absolute number of attendance is highly correlated with the level the games are played. To reduce the chances of multicollinearity the variable attendance is taken as a difference to the mean. These absolute attendance levels could, therefore, be one of the major drivers of the found results regarding the level dummy variables.

Moving on to the second hypotheses the theoretical framework of the research expected that bookmaker's expectations would incorporate the findings found in the former part of this section. As discussed in section four of this research a robust regression analysis is done to calculate the influences the independent variables have on the dependent variable, the expected points by bookmakers.

VARIABLES	(8) Actual Points	(9) Expected points
Distance (100km)	0.0181***	0.00335***
	(0.00678)	(0.000803)
Attendance (1000x)	0.00916***	0.00351***
	(0.00140)	(0.000158)
Last Year Rank	0.00483***	0.00417***
	(0.000786)	(9.34e-05)
Current Rank	0.0142***	0.0174***
	(0.00114)	(0.000135)
Market Value	1.89e-09***	1.74e-09***
	(1.43e-10)	(0)
Form 5 matches	0.00733***	0.00814***
	(0.00249)	(0.000296)
Championship	-0.0808***	-0.0198***
	(0.0260)	(0.00303)
League One	-0.0964***	-0.0277***
	(0.0261)	(0.00303)
League Two	-0.150***	-0.0397***
	(0.0260)	(0.00303)
Constant cut1	-0.627***	
	(0.0263)	
Constant cut2	0.108***	
	(0.0261)	
Constant		1.607***
		(0.00299)
Observations	18,324	18,324
R-squared		0.817
Pseudo R2	0.0290	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3

Table 3 shows the full ordered probit model of the first hypothesis and the robust regression analysis side by side. Without looking at the magnitude of the effects both models seem quite similar. All variables used in this research shown similar signs and degrees of significance. This means that the independent variables play a significant role in both the expectations of the bookmakers as do they in the actual results. To check whether these effects are similar in magnitude as well Table 4 compares the marginal effects of the ordered probit model with the robust regression analysis of the bookmakers' predictions.

VARIABLES	Actual Points	Expected Points	Expected Points [std. err]	Confidence interval [99%]	
Distance (100km)	0.0203658	0.0033543	0.0008032	0.000945	0.005764
Attendance (1000x)	0.0103128	0.0035097	0.0001584	0.003035	0.003985
Last Year Rank	0.0054332	0.0041655	0.0000934	0.003885	0.004446
Current Rank	0.0159809	0.0174155	0.0001348	0.017011	0.01782
Market Value	2.13e-09	1.74e-09	1.59e-11	1.69e-09	1.79e-09
Form 5 matches	0.0082536	0.0081438	0.0002956	0.007257	0.009031
Championship	-0.0910452	-0.0197575	0.0030328	-0.02866	-0.01066
League One	-0.1085991	-0.0277059	0.0030348	-0.03681	-0.0186
League Two	-0.1689498	-0.0397487	0.003034	-0.04885	-0.03065

Table 4

Looking at the results of the robust regression analysis of the bookmaker's expectations we find that the distance traveled by the away team has a significant and positive effect of the expected number of points by the home team. An increase in 100 kilometers travel distance will increase the expected points of the home team by 0.003 points, c.p. Comparing this to the results found when looking at the actual results [0.020 points] we find that the magnitude of these effects are different. The actual effect of traveled distance is six times [15] as big as the effect expected by bookmakers and incorporated in their odds. When looking at the standard

[15] 0.0203658 / 0.0033543 = 6.07155

errors of the coefficients of the robust regression analysis we can conclude that both effects are significate different at the 1% level.

These results suggest that arbitrage is possible due to 'incorrect' incorporation of the effect of traveled distance by the bookmakers. A betting strategy in which the gambler carefully picks his matches and chooses only those matches in which away teams have to travel huge distances will potentially yield profits due to favorable odds. This is in line with the forecast models and conclusions from Goddard and Asimakopoulos who expected an 8 percent margin over money wagered (Goddard & Asimakopoulos, 2004). Therefore we reject the hypothesis set in the theoretical framework of this research that arbitrage is not possible.

Looking at the other variables used in this research the difference between actual and expected effects is smaller than the results found for the variable traveled distance. But an optimal betting strategy would include an emphasis on higher attendance matches (compared to league average), the bigger difference in last year ranking and positive difference in market value. Contrary to this is the effect of the difference in current ranking, for this, the bookmakers over adjust their expectations.

6. Conclusion

This section concludes this research. First, an overview of this research is given followed by a summary of the main findings. These results will be put into perspective given the available literature and implications of these results will be summarized. This section will end with discussing the limitations of this research as well as providing suggestions for further research.

The home team advantage is one of the most interesting phenomena in sports. This advantage is found to be present in all sports disciplines and at all levels. Literature suggests different sources of this home team advantage. In this research, the main source of interest is the traveled distance of the away team. This research has looked at 26,468 games played at the four highest levels of professional English football. These games were played in a time span between 2002 and 2014. During that time period, the home team advantage was 58.6% which is in line with previous findings.

Using an ordered probit model we showed that travel distance has a positive and significant effect at the 1% level. An increase of 100 kilometers traveled distance by the away team increased the home team points by 0.02 points. We, therefore, reject the hypothesis that traveled distance has no effect on match outcomes. To the best of our knowledge, this is the first attempt to calculate the direct effect traveled distance has on the expected number of points. But these findings are in line with previous research who showed a positive relationship between traveled distance and number of goals scored. To answer the main question of this research, “does distance play a role?” a very short answer is sufficient; yes it does. As in standard economics, the physical location does affect the outcome.

What do these results imply? The world football organization FIFA and her European sister UEFA are currently looking to improve the game of football and improve fairness. In that light neutral grounds for finals is maybe not the best solution. A ‘double final’ in which both teams play one of the games at home would be fairer. Both teams would travel the same distance then, instead of a neutral ground which is maybe near one team but not for the other. For instance, the final of the FA CUP is played at Wembly but this favors clubs like Arsenal over Manchester United.

The novelty of this research is that bookmakers' expectations are looked into, to investigate whether these findings regarding the traveled distance are incorporated in their odds. We find that arbitrage is possible due to an incorrect incorporation of the effect traveled distance has on the actual result. Actual effects are six times larger than the incorporated expected effect by bookmakers, concerning traveled distance. This opens up the possibility for profitable betting strategies. Consequently, we rejected the hypothesis that there is no possibility for arbitrage regarding traveled distance.

As this research sole focusses on the four highest divisions of the English football league, conclusions drawn from this research are maybe not true for other countries. But assuming that the English football league is one of the most developed ones, travel distance would have an even bigger effect in countries where transportation is less convenient. The biggest limitation of this research is that the ordered probit model only explains 3 percent of the variance in the dependent variable. Although football is a high game of chance there is a possibility that some variables which do affect the match outcome are unintentionally omitted in this research, leading to an omitted variable bias. Another major limitation of this research is that with using an ordered probit model, heteroscedastic errors could bias the estimator parameters. At this moment there is no real test to control for this and homoscedastic errors are assumed in this research.

An interesting suggestion for further research could be incorporating the results found in this research into the forecasting models to better predict the outcomes of football matches in the future. Bookmakers could then use these predictions to offer the correct odds, without a possibility to outperform them.

Bibliography

Blake, P. (2016, January 16). *Take the long road? The effects of distance travelled on results*. Retrieved from Mathematically Safe: https://mathematicallysafe.wordpress.com/2016/01/16/take-the-long-road-the-effects-of-distance-travelled-on-results/

Boyko, R., Boyko, A., & Boyko, M. (2007). Referee bias contributes to home advantage in English Premiership football. *Journal of Sports Sciences*, 1185-1194.

Cain, M., Law, D., & Peel, D. (2000). The favrourite-longshot bias and market efficieny in UK football betting. *Scottish Journal of Political Economy*, 25-36.

Carmichael, F., & Thomas, D. (2005). Home-Field Effet and Team Performance. *Journal of Sports Economics*, 264-281.

Clarke, S., & Norman, J. (1995). Home grund advantage of individual clubs in English soccer. *The Statistican*, 509-521.

Clay, D., Clay, N., & Bro, A. (2014, February 24). *Hoe Does Travel Affect NCAA Bassketball Outcomes?* Retrieved from Winthrop Itelligence: https://www.canr.msu.edu/csus/uploads/458/49149/How_Does_Travel_Affect_NCAA_Basketball_Outcomes__-_Winthrop_Intelligence.pdf

Courneya, K., & Carron, A. (1992). The home advantage in sport competitions: A literature review. *Journal of Sport and Exercise Psychology*, 13-27.

Deloitte. (2017). *Deloitte Football Money League 2017.* Deloitte.

Dixon, M., & Coles, S. (1997). Modelling Association Football Scores and. *Journal of the Royal Statistical Society*, 265-280.

Edwards, J. (1979). The home field advantage. In J. Goldstein, *Sports, games, and play: Social and psychological viewpoints* (pp. 409-438). Hillsdale: NJ: Lawrence Erlbaum.

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 383-417.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 331-340.

Goddard, J. (2006). Who wind the football. *Significance*, 16-19.

Goddard, J., & Asimakopoulos, I. (2004). Forecasting Football Results and the Efficience of Fixed-odds Betting. *Journal of Forecasting*, 51-66.

Gratton, C. (1998). The Economic Importance of Modern Sport. *Culture, Sport, Society*, 101-117.

Kent, B. (2015, May 20). *Does Tracel Distance Affect Results in MLS?* Retrieved from American Soccer Analysis: http://www.americansocceranalysis.com/home/2015/5/20/zrq8500mrlga9pnts3basp37l2yi7r

Liao, T. (1994). *Interpreting Probability Models; Logit, Probit, and Other Generalized Linear Models.* SAGE Publications, Inc.

Limao, N., & Venables, A. (2001). Infrastructure, Geographical Disadvantage, Transport Costs, and Trade. *The World Bank Economic Review*, 451-479.

Nevill, A., Newell, S., & Gale, S. (1996). Factors associated with home advantage in English and Scottish soccer matches. *Journal of Sports Sciences*, 181-186.

Pollard, R. (1986). Home advantage in soccer: a retrospective analysis. *Journal of Sport Sciences*, 237-248.

Pollard, R. (2006). Worldwide region variations in home advantage in association football. *Journal of Sport Sciences*, 231-240.

Pollard, R. (2008). Home Advantage in Football: A current review of an Unsolved Puzzle. *The Open Sports Sciences Journal*, 12-14.

Pollard, R., & Seckin, A. (2007). Why is home advantage in South-east Europe the highest in the world? *12th European Congress of Sport Psychology*, 53-56.

Pollard, R., Silva, C., & Medeiros, N. (2008). Home advantage in football in Bazil: differnces between teams and the effects of distance traveled. *The brazilian journal of soccer science*, 3-10.

Schechtman-Rook, A. (2013, May). *Home Field Advantage: Distance Doesn't Matter, But Time Zones Do*. Retrieved from An Ivory Tower Take on NFL Analytics: http://phdfootball.blogspot.nl/2013/05/home-field-advantage-distance-doesnt.html

Schwartz, B., & Barsky, S. (1977). The Home Advantage. *Social Forces*, 641-661.

Snyder, E., & Purdy, D. (1985). The home advantage in collegiate basketball. *Sociology of sport journal*, 352-356.

Sporza.be. (2017, May 29). *Chabarovsk-fans staan voor uitdaging: dagen reizen voor uitwedstrijden*. Retrieved from Sporza.be: http://sporza.be/cm/sporza/extra/opvallend/1.2992021

Stefani, R., & Clarke, S. (1992). Predictions and home afvantage for Australian rules Football. *Journal of Applied Statistics*, 251-261.

Swartz, T. B., & Arce, A. (2014). New Insights Involving the Home Team Advantage. *International Journal of Sports Science & Coaching*, 681-692.

Thaler, R., & Ziemba, W. (1988). Parimutel Betting Markets; Racetrack and Lotteries. *Journal of Economic Perspectives*, 161-174.

Williams, R. (2008). Ordinal regression models: Problems, solutions, and problems with the solutions. *German Stata User Group Meetings*.

Appendix

Appendix A

AFC Bournemouth
AFC Wimbledon
Accrington Stanley FC
Aldershot Town FC
Arsenal FC
Aston Villa FC
Barnet FC
Barnsley FC
Birmingham City FC
Blackburn Rovers FC
Blackpool FC
Bolton Wanderers FC
Boston United FC
Bradford City AFC
Brentford FC
Brighton & Hove Albion AFC
Bristol City FC
Bristol Rovers FC
Burnley FC
Burton Albion
Bury
Cambridge United FC
Cardiff City FC
Carlisle United FC
Charlton Athletic FC
Chelsea FC
Cheltenham Town FC
Chester FC
Chesterfield FC
Colchester United FC
Coventry City FC
Crawley Town FC
Crewe Alexandra FC
Crystal Palace FC
Dagenham & Redbridge FC
Darlington FC
Derby County FC
Doncaster Rovers FC
Everton FC
Exeter City FC
Fleetwood Town FC
Fulham FC
Gillingham FC
Grimsby Town FC
Halifax Town FC
Hartlepool United FC
Hereford United FC
Huddersfield Town AFC
Hull City AFC
Ipswich Town FC
Kidderminster Harriers FC
Leeds United FC
Leicester City FC
Leyton Orient FC
Lincoln City FC
Liverpool FC
Luton Town FC
Macclesfield Town FC
Manchester City FC
Manchester United
Mansfield Town FC
Middlesbrough FC
Millwall FC
Milton Keynes Dons FC
Morecambe FC
Newcastle United FC
Newport County AFC
Northampton Town FC
Norwich City FC
Nottingham Forest FC
Notts County FC
Oldham Athletic AFC
Oxford United FC
Peterborough United FC
Plymouth Argyle FC
Port Vale FC
Portsmouth FC
Preston North End FC
Queens Park Rangers
Reading FC
Rochdale AFC
Rotherham United FC
Rushden & Diamonds FC
Scunthorpe United FC
Sheffield United FC
Sheffield Wednesday FC
Shrewsbury Town FC
Southampton FC
Southend United FC
Stevenage FC
Stockport County FC
Stoke City FC
Sunderland AFC
Swansea City AFC
Swindon Town FC
Torquay United FC
Tottenham Hotspur FC
Tranmere Rovers
Walsall FC
Watford FC
West Bromwich Albion FC
West Ham United FC
Wigan Athletic FC
Wimbledon
Wolverhampton Wanderers FC
Wrexham AFC
Wycombe Wanderers FC
Yeovil Town FC
York City FC

Appendix B

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Home Points	1											
(2) Expected Home Points	0.27	1										
(3) Distance (100km)	0.02	0.01	1									
(4) Attendance (1000x)	0.14	0.44	-0.02	1								
(5) Last Year Rank	0.13	0.45	0	0.26	1							
(6) Current Rank	0.18	0.75	-0.01	0.25	0.25	1						
(7) Market value	0.18	0.65	0	0.45	0.34	0.27	1					
(8) Form 3 Matches	0.11	0.46	0	0.15	0.15	0.49	0.2	1				
(9) Form 5 Matches	0.13	0.57	-0.01	0.2	0.18	0.59	0.25	0.81	1			
(10) Championship	0	0.01	-0.03	0	0	0	0	0	-0.01	1		
(11) League One	0	0	0.03	0	0	0	0	0	0	-0.37	1	
(12) League Two	-0.02	-0.02	0	0	0	0	0	0	0.01	-0.37	-0.37	1